

# Data Mining Approach to Implement a Recommendation System for Electronic Tour Guides

Mohamad Saraee and Shafiullah Khan  
School of Computing, Science and Engineering  
University of Salford, M5 4WT

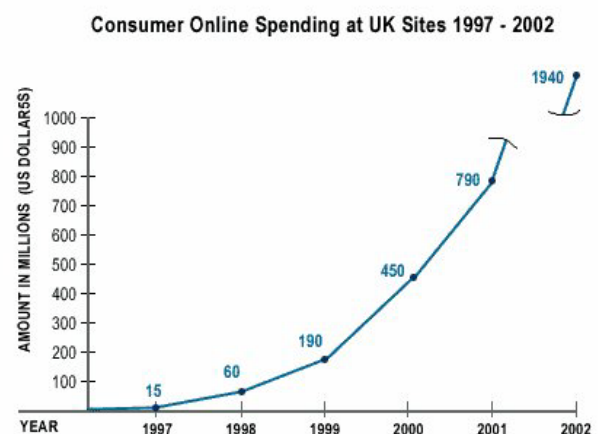
## Abstract

In this paper we consider the problem of discovering patterns generated by association rules between items in a database of locations that tourists have visited within the Manchester city centre area. The focus of the paper is on the interactive agenda offered by a tourist information system. This assist the user in the organisation of a visit or a tour by providing personalised recommendations dependent upon the calculations made by the association rule algorithm using the entries within a database of tourist attractions. The paper shows the development of a stand alone recommendation system, which uses association rule algorithm to propose and recommend any related locations that a particular tourist can visit based on the option they initially choose. This opens the door to take this implementation further to an extent where developing this system on a manufacturing scale is viable and be installed in major cities around the UK so that tourists can truly appreciate their time in cities like Manchester. Before such mass production can take place other data mining techniques such as clustering or temporal data mining are suitable extensions to such an application. This can offer further interactive options to tourists who use the application as a means of visiting major attractions and fulfilling their holidays.

## 1 Introduction

Association Rule is a powerful data analysis technique that appears often within the data mining subject area. The analysis of association rule is used in a variety of ways and in many applications, including

merchandise stocking, insurance fraud investigations, medical diagnosis and even climate prediction. However there is a distinct lack of growth of data mining applications within the tourism industry. Recommendation systems are commonly defined as applications that e-commerce sites exploit to suggest products and provide users with information to facilitate their decision-making processes. The graph in figure 1 shows the rise in consumer spending at UK sites. This provides a good indication to the commercial strength of such sites and the systems they employ.



Source: Datamonitor

Figure 1: Consumer spending at UK sites.

They implicitly assume that they can cater for user needs and constraints, through appropriate recommendation algorithms, and convert them into product selections using the knowledge compiled into the system. This then expects the users to remember or make a note of the information provided by such applications. The system is designed to tackle many of these problems.

1. This paper presents and highlights data mining and in particular a novel Association Rule algorithm, which enables the opportunity of development of more conveniently available applications, which are stand-alone and are based in carefully, selected zones such as major train stations for example. The algorithm used can provide a service for tourists on the basis of past tourist experiences and their choices. It will however leave the focus on the user's choice rather than recommending many attractions and going off on a tangent from the user's original choice. By combining this algorithm with temporal data mining techniques it can provide state-of-the-art data mining application that can then open major commercial opportunities and take tourism to another level.
2. The system will be based upon domain-intrinsic information and as a result provide a competent and practical service to the tourists who visit the city of Manchester. The main contributions of this paper are as follows: (1) Association Rule mining within the tourism domain. (2) An attempt towards the development of an interactive agenda towards the interface to provide the required information. (3) To provide a stand-alone system, which can be provide information on local attractions nearer to the destination rather than online. (4) To open a well thought through commercial opportunity via the availability of the recommendation system within major cities.
3. The remainder of this paper is organised as follows. In section 2 we review data mining techniques and in particular Association Rule mining and what other techniques can be built

upon this for further development. In section 3 we look at Association Rule mining in more depth the theory behind it, and what it can produce within an application used in the tourism domain. Section 4 is based upon the development of the system and how it can be used within the real world and finally, some conclusions are drawn in Section 5.

## 2 Problem Statement

The implementation that this paper relates to is very much an application oriented project. It takes into account a domain, which is important to any big city, the domain being the tourism industry. The major problems with current applications providing information for tourists is, they are all web related and can only be accessed online via a website based system. There are some obvious advantages for this type of an application, noticeably having 24hr access from the comfort of ones home. However it is a major setback for a tourist in the hustle and bustle of a major English city to forget some vital information about an attraction, whether it be directions, opening times, pricing information or even related attractions that may be of interest to them. Therefore this project opens the door for further technical opportunities as well commercial scope. The recommendation system for electronic tour guides will provide tourists with the relevant information about the attraction that they are interested in. Not only that, it will also provide recommendations so that tourists can make full use of their tours of Manchester. From a commercial perspective, the system will be stand-alone pieces of equipment, which means that it can be installed wherever it is appropriate, such as major train stations, town squares and even major shopping malls. Together with the commercial growth it will also provide users to access relevant information when it is suitable for them, supplying the flexibility to access the information once in the city

centre and not only online via a website on a computer. From a technical viewpoint, the application will be based upon the Association Rule mining analysis and will provide recommendations based on previous tourists' behaviour i.e. from a database. However there still remains the scope to further the data mining analysis for such an application such as temporal analysis and classification analysis. Nevertheless Association Rule mining provides a strong basis as it is based upon the notion of frequent patterns, associations and correlations within databases and to detect these interesting and useful patterns.

## 2.1 Related Work

There have been many WWW based applications that are developed in years gone by, such recommendation systems as VacationCoach and TripleHop are mainly implemented as a means of assisting the user's decision making process and a method of exploiting the commercial opportunity. Such data mining applications have various ways in which they conduct their respected technical awareness and ability with some using user profiling methods and others using estimations. The VacationCoach and TripleHop approaches can be found in [1]. Previous work has had a tendency to put the onus on the WWW to create a niche for tourism systems. However, online encounters of tourism information are currently restricted to interactions with interfaces that more or less directly mirror the ontology of the database systems to which they are connected [2]. This again opens the door to provide a more interactive approach to the problem, supplying the users to access the information when they want with a greater sense of mobility that does not cost anything for the user. It is an approach that can bridge the gap between the commercially driven WWW based tourist systems and the more expensive, poorly available yet highly portable mobile devices with applications

for electronic personal tour guides; so we can see clearly where this is heading.

## 3. Methodology

During the design of the application the main technique used for data mining purposes is the Association Rule, which looks for the pattern in terms of its frequency and the relationship between other itemsets. In terms of the full-blown application and the interactive agenda of it, it is very useful to have the ability to provide the information necessary to the user. However, it is equally if not more important to allow the application to recommend other attractions within the Manchester city centre area, so that it offers other alternate options for users to visit on their tour of the city. This will be based upon a large dataset, which holds data about the attractions previous tourists visited on their tours of the city. There are data regarding to many attractions however for the purpose of this work 15 attractions compose the dataset. Naturally, not all tourists visit all 15 attractions and therefore Association Rules mining technique is particularly good for this distinctive scenario, as it takes into account the support of those transactions that have taken place (visiting) and even those that have not taken place (not visiting). A diagrammatic description can be seen in figure 3.

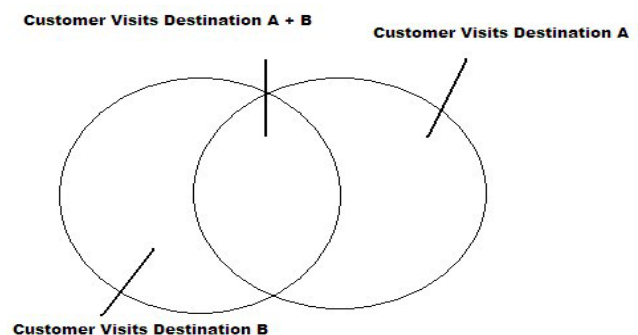


Figure 3: Customer Visits

Using this particular technique will provide the necessary information to the user as well accommodate the flexibility of building

further on the existing technique. Such further building may include techniques such as temporal mining and even classification techniques, resulting in more options available for the user and a more hi-tech application.

## **4. Experimental Evaluation**

After integrating the Association Rule mining techniques together with the interface, it provides a simple yet effective core on which further techniques can be built, nevertheless on its own it caters for a simple interface architecture that provides the necessary and relevant information to the user. It will also provide an uncomplicated approach for the user to utilise the recommendations that the application provides for the user to manage their tour most effectively.

### **4.1 Experimental Setup**

If this application was to be installed in the real world, the process of using the system will not involve the user to interact with system via peripheral devices such as a mouse. The user will be able to access the information via a touch screen, this is much more practical to install within public places around the city of Manchester. However for the means of this setup, it is vital that the system can offer the services to various types of users (i.e. tourists) by using peripheral devices. The application will simply be switched on and ready to use all the time, so that in a real life situation it should not matter when a tourist approaches the system and digest the information provided by the system and plan their tour to their liking. The operational aspects of the system are very commercially biased however the technical side of the system should be able to handle the transactions (i.e. user requests) at any time of day. This should not be a problem due to the nature of the software abilities and accessibility of the application.

## **4.2 Discussion**

The prospects for higher-level recommendation software applications in this tourist information system domain are immense.

The best market opportunity for such implementations will probably be the interoperation of such systems to alleviate the integration of legacy tourist information systems within modern environments which would provide tourists with greater flexibility to obtain the information when it is needed the most.

The overall industry itself and the domain, within which these techniques are used, is confronted with interfacing difficulties. This is due to the reason the systems being very heterogeneous (a multitude of underlying operating systems, communication protocols, etc.), dynamic (reservation status may for example change on very a short time basis) and conceived for diverse contexts (travel agencies, airlines, hotel reservations, etc.). As a result a typical WWW based recommendation system will be based upon multiple contexts within it can operate and cater for every aspect of tourists' visits. This is where the system can come in and assist a tourist, not in conjunction with any travel agencies or hotel reservation but be solely implemented on the basis of providing information to a tourist when it is needed the most, when newly arriving to a major city like Manchester.

## **5. Conclusion**

Overall the recommendation system is capable of displaying information about tourist attractions within a particular city. It can display information such as what is available at the local yet renowned cinemas or even what concert is playing this week at the Theatres around the city centre. To get the best use out of the system it is ideal for it to be installed and displayed at specific locations such as train stations and airports. This is because these are naturally the most

likely place where tourists start their journeys and tours of major cities. The system does have some limitations but these can be transformed by using other data mining techniques such as temporal mining and classification mining. One of the limits that can be identified at this moment in time is the lack of spatial awareness, which can provide information about theatres starting with the closest to that point in terms of distance. The main reason behind this is due to the fact that most tourists would use public transport for means of travel. Also, the system is based on the exploitation of specialised hardware, physically located in the place to be visited, and on the use of special hardware devices such as touch screen and specialised security casing around the application.

Future work that can be useful to the system may well involve deeper mining techniques like mentioned above. Temporal data mining could be used to see what is on offer in the city based on the season as many attractions are naturally based on seasonal trade i.e. the system can show season attractions for this summer. Temporal mining is a great technique as it gives you the accuracy and precision required for this system. Another way of improving the system is using classification i.e. breaking the system into a hierarchical structure as shown in Figure 5.

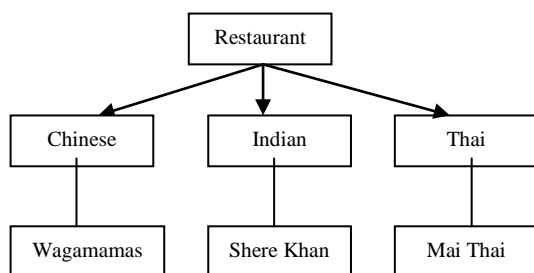


Figure 5: Hierarchical Structure

Not all tourists prefer the same attractions, a brief example below shows restaurant being the preferred option by a typical tourist. It then classifies the types of restaurants available (from the database) it would provide the options which are much more

specific giving the user a more defined result and recommendation.

## 6. References

### Papers

- [1] Francesco Ricci, Travel Recommender Systems, IEEE Intelligent Systems, Nov./Dec. 2002, pp. 55-57
- [2] Alexander Zipf, Adaptive context-aware mobility support for tourists, IEEE Intelligent Systems, Nov./Dec. 2002, pp. 57-59
- [3] Thomas Steiner, Software Agents in Tourism Industry – Prototypes and Prospects, Informatik Intelligent Agents, No.1 Feb 2000
- [4] Liliana Ardissono, Anna Goy et al., Ubiquitous User Assistance in a Tourist Information Server, University of Torino, Torino, Italy
- [5] <http://informatics.indiana.edu>
- [6] [http://www.sims.berkeley/~kirsten/rec\\_experiment.htm](http://www.sims.berkeley/~kirsten/rec_experiment.htm)